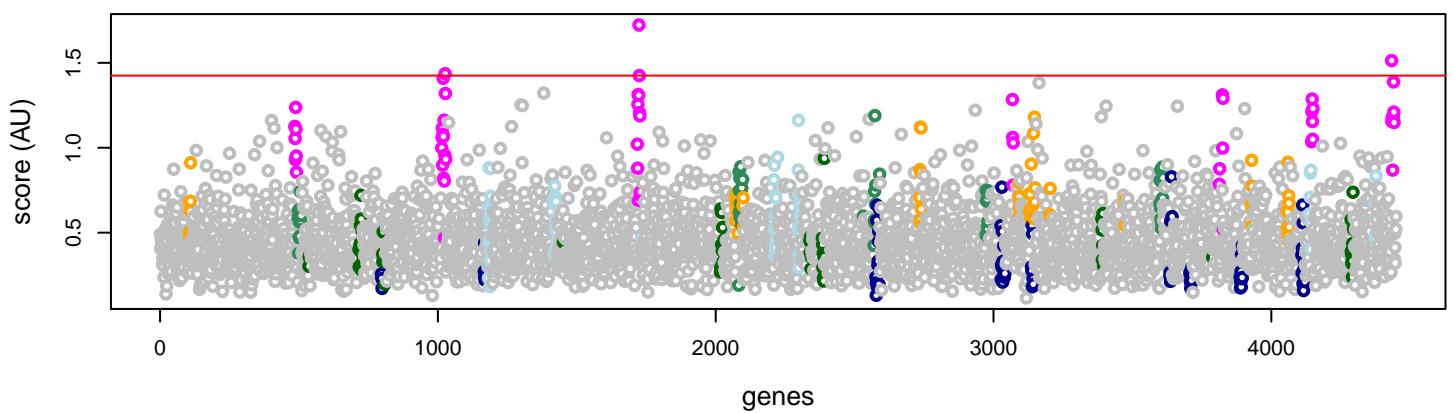
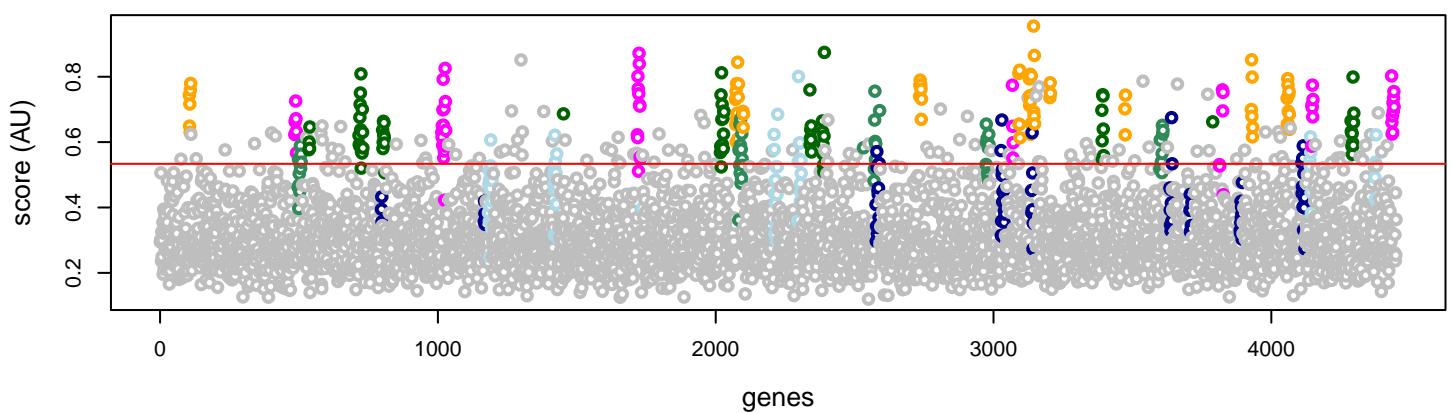
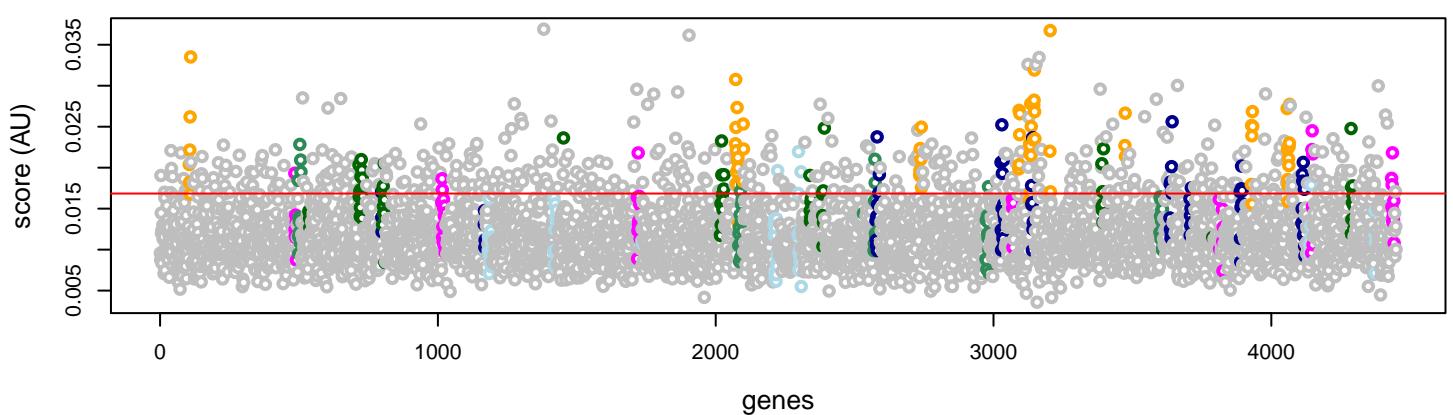
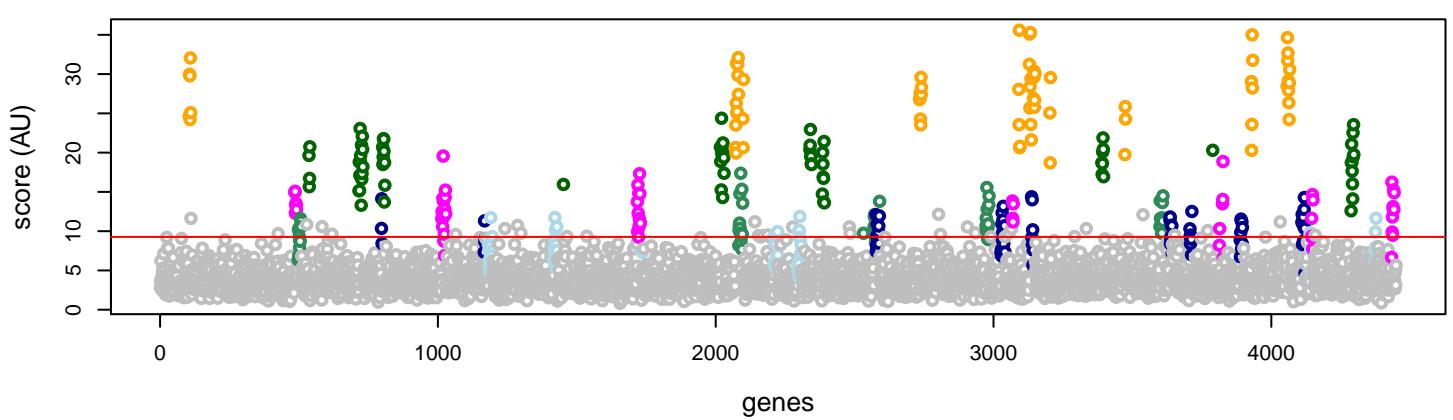
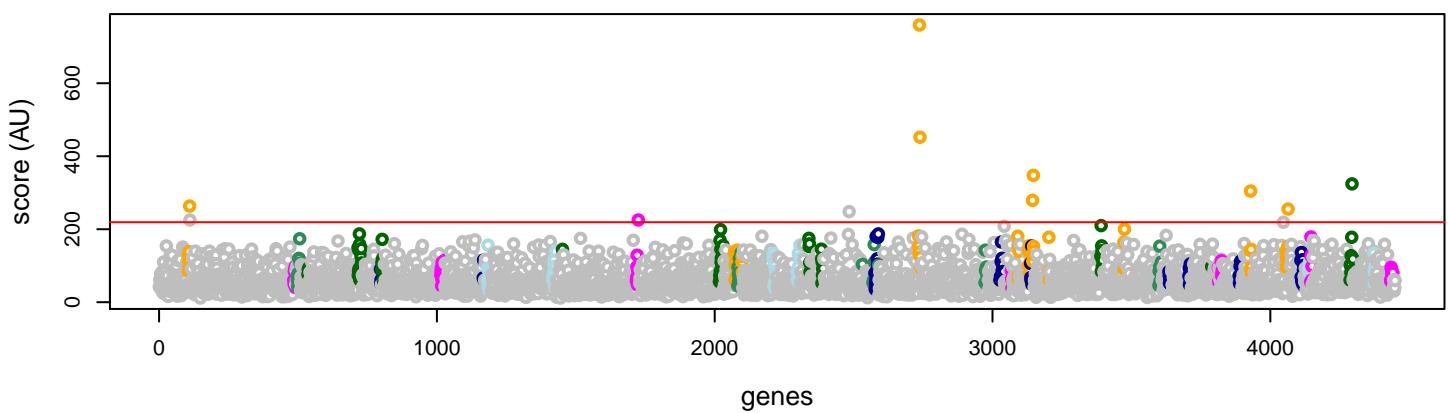


**Figure S1: Atypicality score of the genes along one of the “standard” model genome for each of the 16 methods.**

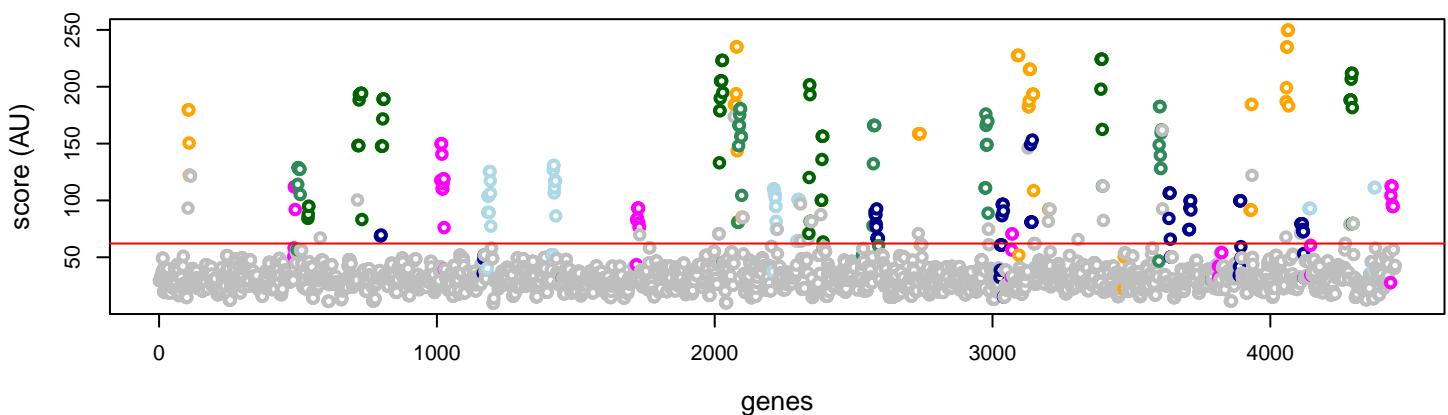
Each dot corresponds to the score value of a gene. The color of the dot represents the origin of the gene: native genes (*i.e.* *E. coli* genes) are in grey; HTs are in color according to Table 1 (shades of blue are for close genomes, shades of green for intermediary genomes and pink and orange are for far genomes, see M&M). The horizontal red line corresponds to the optimal threshold value (*i.e.* the one that minimizes mean error of detection, see M&M). (See figures on following pages)

**CU.chi2****CU.karlin****CU.karlin.aa****CU.KL**

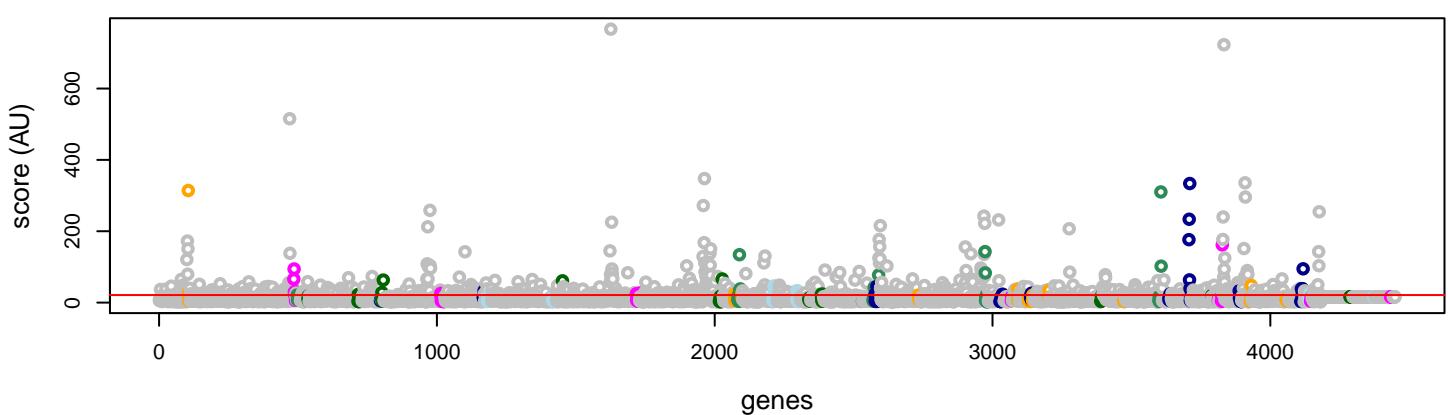
### CU.mahalanobis



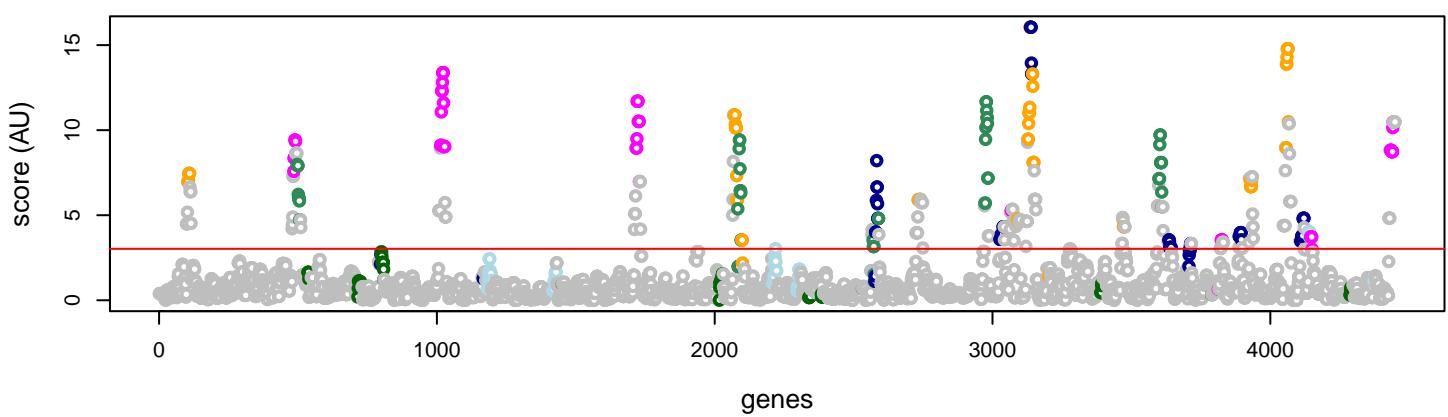
### dint5

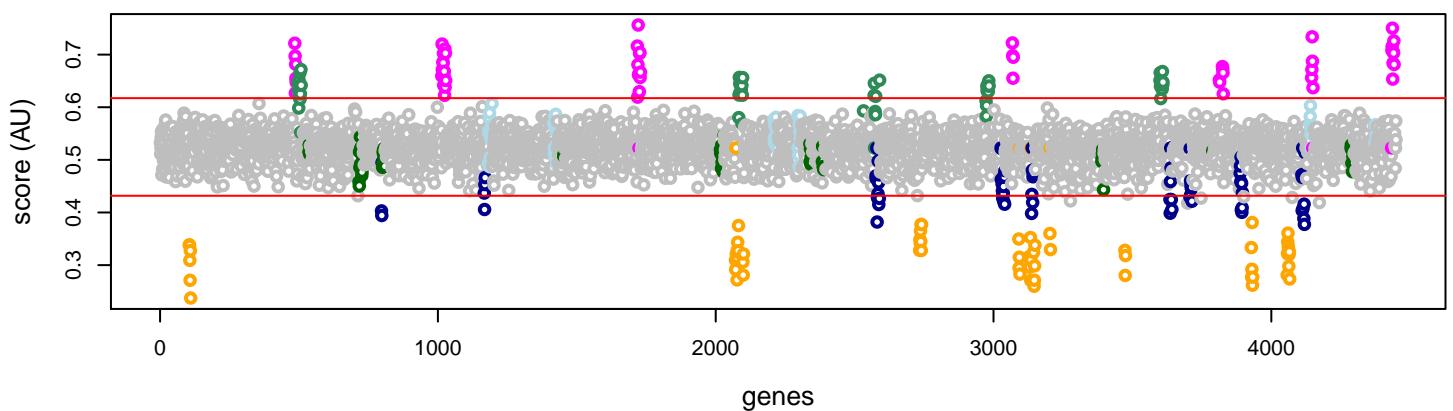
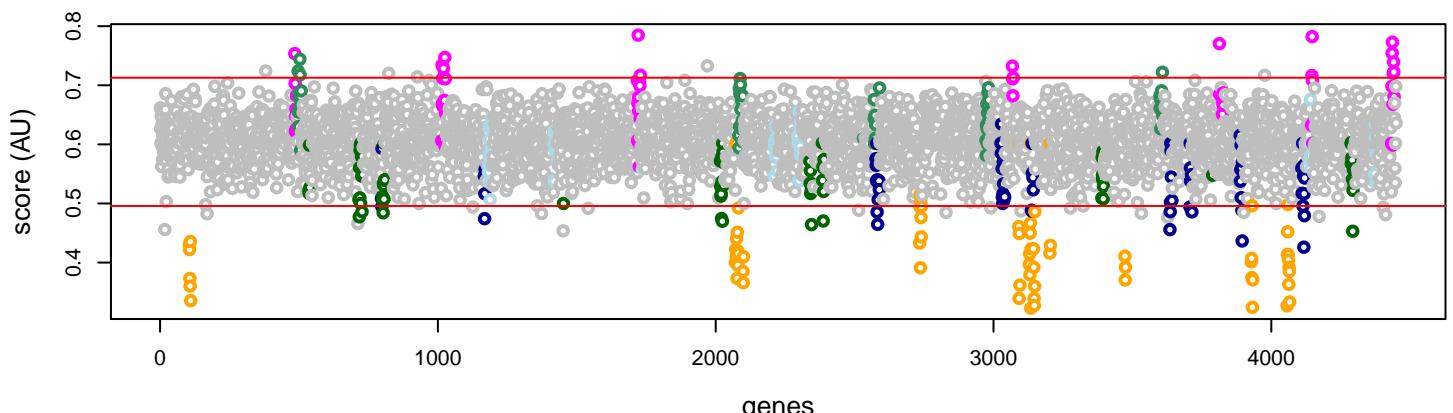
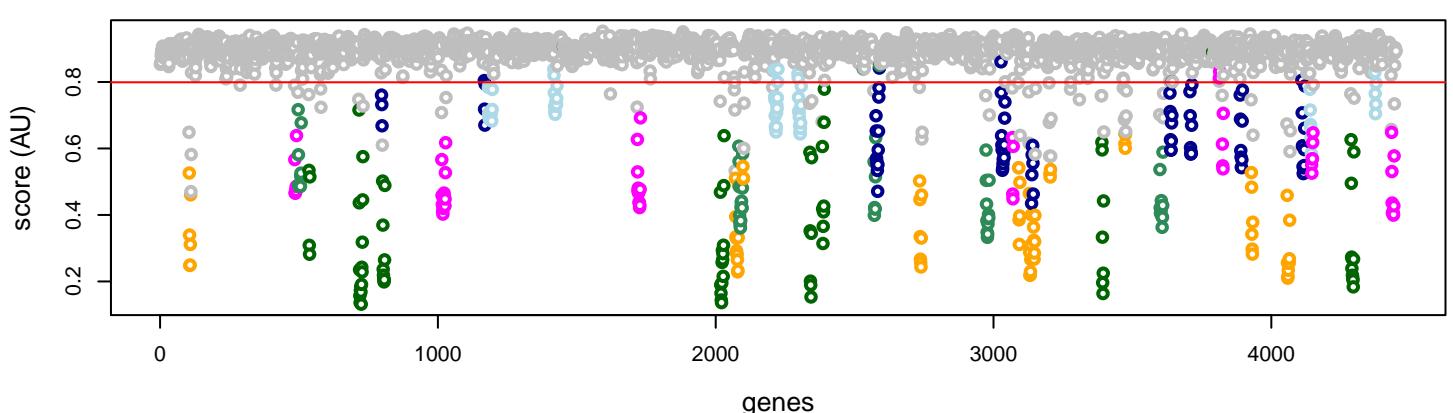
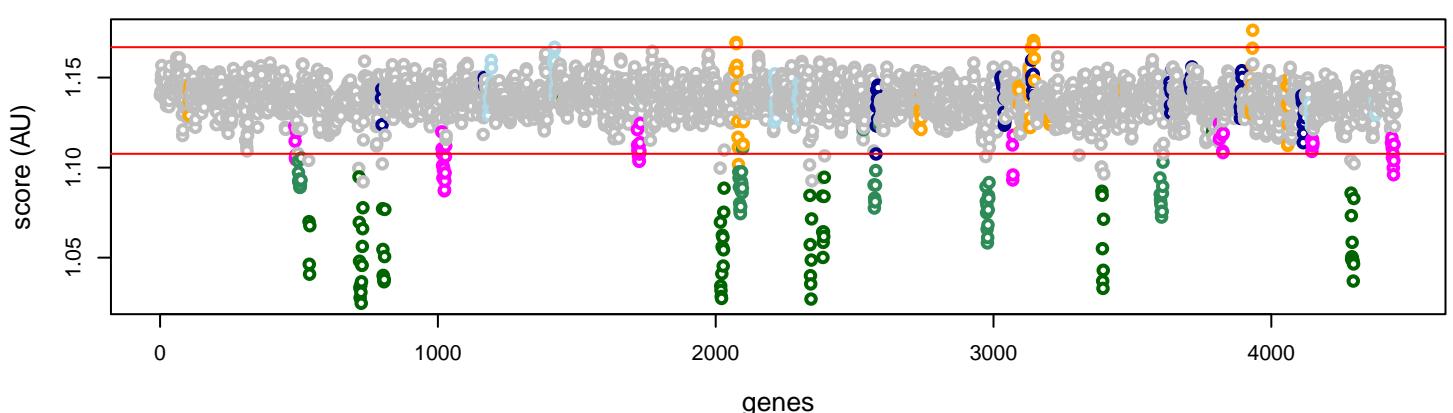


### dint.di31T2

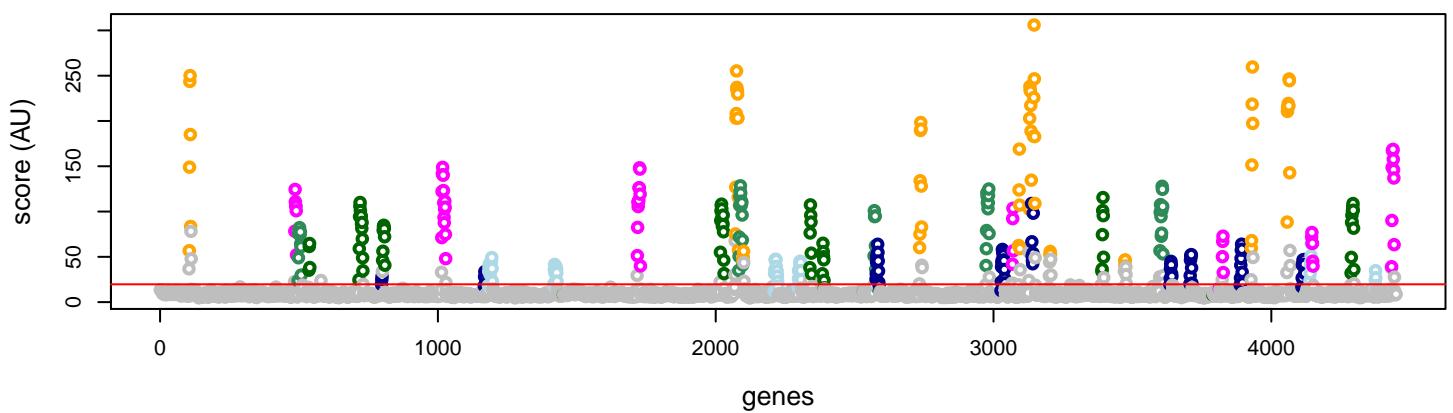


### GC.windows

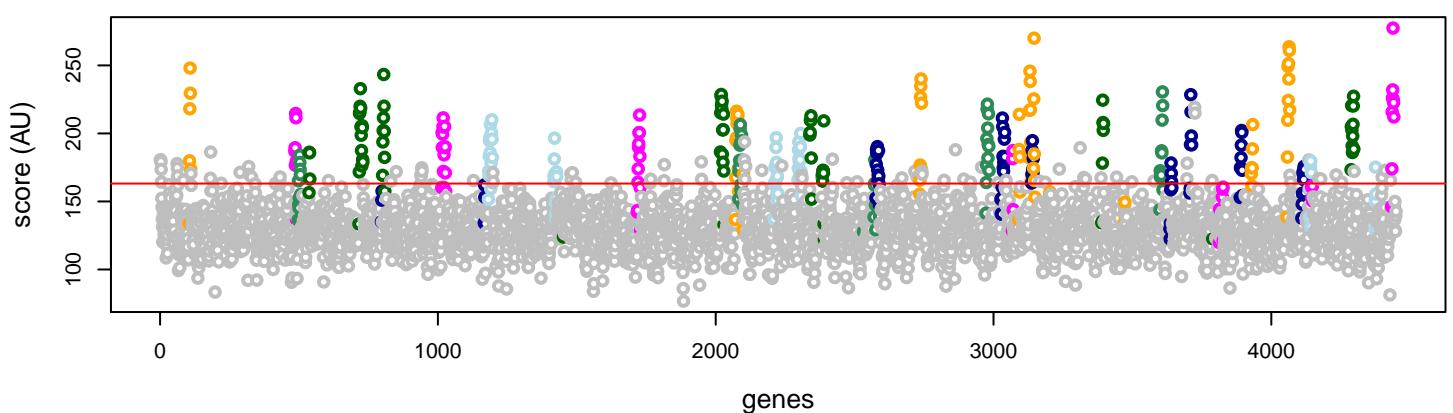


**GCtotal****GC1-GC3****oli.Pearson****oli.covariance**

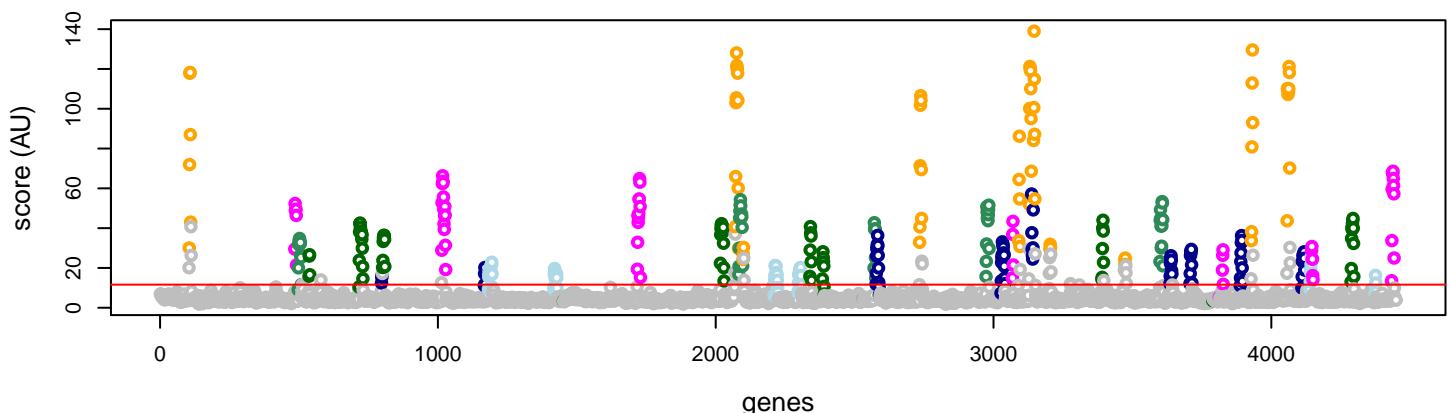
**oli.chi2**



**oli.mahalanobis**



**oli.KL**



**signature**

